Initial Selection of a GCOS Surface Network

Thomas Peterson,* Harald Daan,+ and Philip Jones#

ABSTRACT

To monitor the world's climate adequately, scientists need data from the "best" climate stations exchanged internationally on a real-time basis. To make this vision a reality, a global surface reference climatological station network is in the process of being established through the Global Climate Observing System (GCOS). To initially select stations to be considered for inclusion in this GCOS Surface Network, a methodology was developed to rank and compare land surface weather observing stations from around the world from a climate perspective and then select the best stations in each region that would create an evenly distributed network. This initial selection process laid the groundwork for and facilitates the subsequent review by World Meteorological Organization member countries, which will be an important step in establishing the GCOS Surface Network.

1. Introduction

The World Meteorological Organization (WMO) Commission for Climatology (CCI) and Commission for Basic Systems (CBS) are working jointly with the Global Climate Observing System (GCOS) to establish a global reference network of land surface observation stations that would accommodate observed data from most land areas, including many midoceanic islands, at an approximate density of one station per 250 000 square kilometers (World Meteorological Organization 1988a). This density of stations is considered adequate, in combination with representative sea surface temperature data, to monitor global and large hemispheric temperature variability and would permit some multielement analysis, although analysis of elements with lower spatial correlation than temperature (e.g., precipitation) may require denser networks. It is intended that the network be regarded as a standard for developing and improving denser national networks and that the existence of the network will encourage the preservation and exchange of data into the future.

The objectives of the Global Climate Observing System include providing the data required to meet the needs for climate system monitoring, climate change detection, and research toward improved understanding, modeling, and prediction of the climate system (Spence and Townshend 1995). Since satellites cannot provide data needed for long-term—decade to century scale—climate monitoring, this effort focuses on the selection of GCOS land surface observing stations. Other GCOS projects are under way to improve (or limit the degradation to) our observing capabilities in all aspects of the climate system (World Meteorological Organization 1995) ranging from oceans (Nowlin et al. 1996) to the GCOS Upper Air Network (GUAN; World Meteorological Organization 1994).

Currently a large number of weather stations report synoptic or monthly climate (CLIMAT; World Meteorological Organization 1988b) messages over the Global Telecommunication System (GTS) of the WMO. However, these stations may not be the best stations for climate monitoring: many are from urban areas; many are recent stations rather than the very best long-term, homogeneous climate stations needed for climate studies; and their spatial distribution is very

*Global Climate Laboratory, National Climatic Data Center, Asheville, North Carolina.
+Koninklijk Nederlands Meteorologisch Instituut, De Bilt, The Netherlands.
#Climatic Research Unit, University of East Anglia, Norwich, United Kingdom.
Corresponding author address: Thomas C. Peterson, Global Climate Laboratory, National Climatic Data Center, 151 Patton Avenue, Asheville, NC 28801.
E-mail: tpeterso@ncdc.noaa.gov
In final form 17 April 1997.

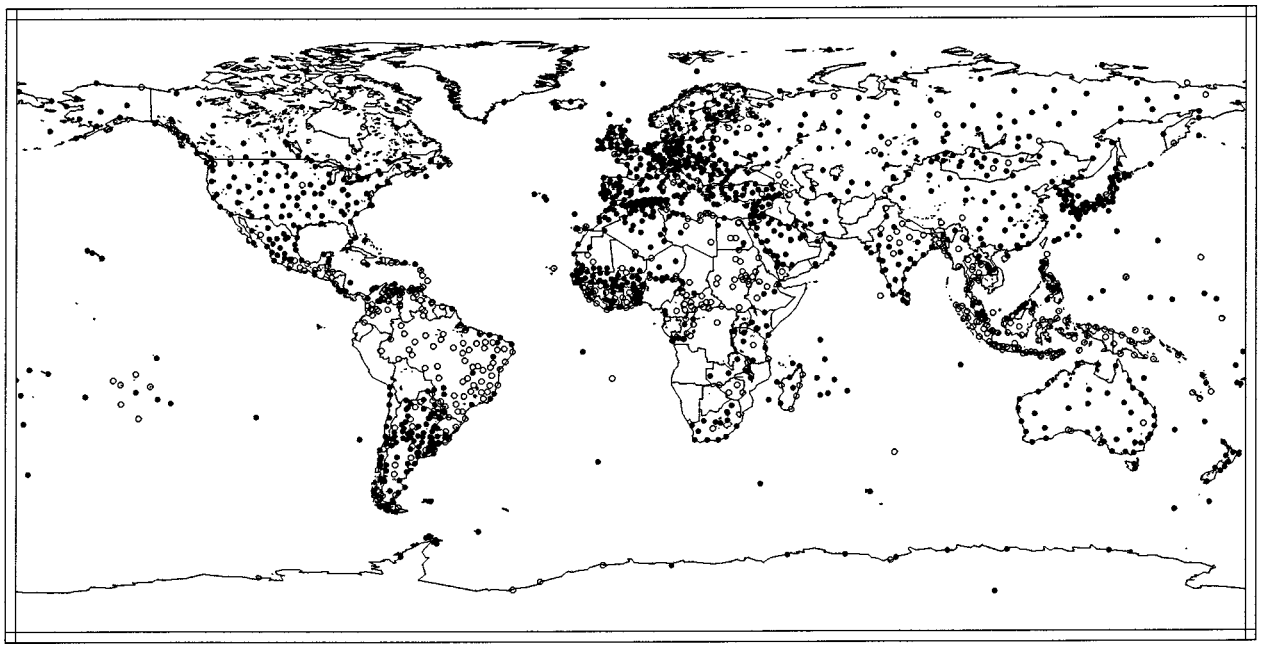

FIG. 1. Stations reporting CLIMAT monthly mean temperature data in 1995. Solid dots (open circles) depict stations with more (less) than 60% of their CLIMAT temperature data received in 1995. While more than 1600 stations reported in 1995, their spatial distribution is uneven.

uneven. As shown in Fig. 1, there are large areas of the world, such as South America from the equator to 20°S, with very sparse CLIMAT reports, while other areas, including, interestingly, sub-Saharan west Africa, have a very dense network of CLIMAT stations. The selection of GCOS Surface Network (GSN) stations, by contrast, needs to be based on the suitability of data for climate analysis resulting in a well distributed network of the very best long-term climate stations in the world. This article describes the procedure that was used to initially select stations for the GSN via a specially developed computer algorithm. WMO members have been informed of this process and asked to review and comment on the selection of stations in their country. How WMO members alter these initial selections may be the subject of a future note.

2. What stations are available for possible inclusion in the GCOS Surface Network?

WMO Publication 9, Volume A (World Meteorological Organization 1996a), lists all weather stations with WMO numbers. Unfortunately, many climate observing stations do not have WMO numbers and Volume A does not indicate whether a station has long enough records to be of use for climate studies. There are, however, several large global datasets used for climate studies. Two of them, the Global Historical Climatology Network (GHCN; Vose et al. 1992; Peterson and Vose 1997), produced in the United States, and P. Jones's dataset (Jones 1994) from the United Kingdom contain (probably) most of the international, long-term, monthly, land surface station temperature data available digitally to researchers. These two datasets were created for slightly different purposes, utilizing different homogeneity testing methodologies, so the information they contain about the stations is not identical. GHCN has 7283 stations and Jones has 2525 stations after removing a subset of U.S. stations that were certain to be duplicates of U.S. GHCN stations. All of the Jones stations are either homogeneous or adjusted to be so, while GHCN has a subset of homogeneity adjusted station data. All data for stations listed in GHCN or Jones are available to researchers at the present time.

Two sources of information at WMO indicate that climate observations are being made at additional stations though the data may not be available for international exchange at the present time. These are the WMO 1961–1990 normals stations and the lists of Reference Climatological Stations (RCSs) that have been submitted by many WMO member states. In May

1996, when the GSN stations were being selected, there were 3342 stations in the normals list, generally having at least 30 yr of data, and the RCS list included 2283 stations with a wide variety of length of records listed according to the response to the WMO enquiry in 1990.

3. Merging the possible stations into a single list

These four sources total 15 433 stations. But how many are unique? To answer this question, all the stations needed to be merged into a single list, as shown in Fig. 2. This was not an easy task: in the four sources, latitudes and longitudes were recorded in hundredths of degrees, tenths of degrees, and minutes, depending on the source. Therefore, station locations seldom agreed exactly. In addition, stations often move slightly over time so data from earlier sources might not have the current location. Not only did the spelling of station names change over time, but in many cases, particularly in newly independent countries, the entire station name changed. These problems are compounded by keypunch errors that occur when keying in hard-to-read metadata from handwritten RCS lists. WMO station numbers also change with time, not only for individual stations. Sometimes a country will change all their WMO numbers at one time. Since most of the metadata flags were keyed to WMO numbers, the final station list was merged with the current WMO Volume A station list to tag each station with the appropriate WMO number whenever possible.

The result was a list of 8653 unique stations, but the list is far from perfect. It is very likely that some errors in merging remain. An example is the assigning of a WMO number from a short-term synoptic reporting airport station to a nearby long-term climate station with a similar name. Also, normals stations were treated as if they started in 1961, which is often erroneous but was the best that could be done with the information available. Therefore, the list should be carefully inspected for accuracy by individuals from each country that has stations initially selected for the GSN.

4. Ranking the stations

The information desired for each station included the length of time the station has been in operation,

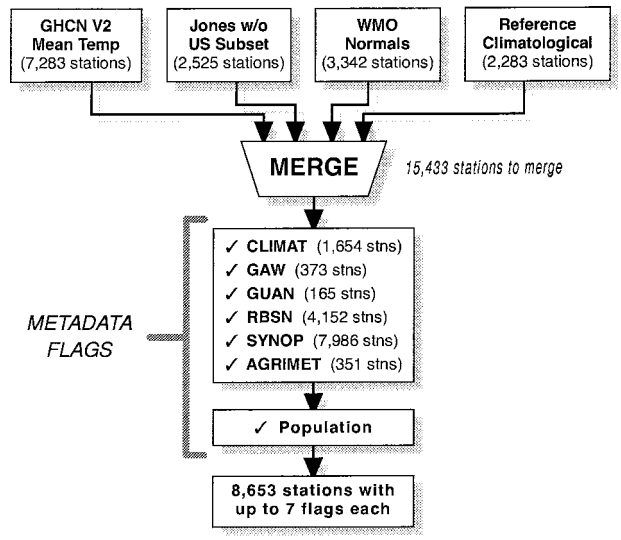

FIG. 2. The process of merging information about potential GCOS Surface Network stations.

the quality of the data, the likelihood of continuing operations, and its ability to report in near-real time. Since most of this information is not available, the problem was approached by examining what metadata could be obtained that would have implications about these aspects. Many of the bits of metadata uncovered had multiple implications.

All 8653 stations were ranked based on an algorithm that resulted from discussions at the Joint CCL–CBS Expert Meeting on the GCOS Surface Network held in Norwich, United Kingdom, in March 1996. For example, approximately how much weight to give to long-term climate records compared to current reporting was agreed upon. All the metadata fall into 10 categories, which are combined in the algorithm to give a total of 100 points.

a. Data

There were 20 points available for (number of years of data)/100.

Since the goal was to select long-term climate stations, the length of data was given considerable weight. For example, a station would receive 10 points if it had 50 years of data and 20 points if it had 100 years of data. However, data prior to 1896 were not considered, so a station with 150 years of continuous operation would also receive 20 points. For GHCN and Jones stations, the numbers used were the actual number of months of data in the last 100 years. For normals stations, data were assumed to be 100% present for the 30 years from 1961 to 1990. RCS stations, some with

very long periods of record indicated, were assumed to have 95% of the data available from the start of their records to the present day. The 95% value is comparable to the average availability in long-term stations in the GHCN and Jones datasets. When a station was listed in more than one source, the longest period of record was used.

b. Homogeneous data

There were 20 points available for (number of years of homogeneous data)/50.

The homogeneity of the data relates to data quality (Easterling et al. 1996). While many stations may be homogeneous, the simplest way to tell if they are indeed homogeneous (or able to be adjusted to make them homogeneous) is if they were included in the Jones dataset or in the GHCN version 2 homogeneity adjusted subset. This data quality aspect is weighted heavily and normalized over the last 50 years. This approach gives weight only to stations whose data are already available internationally.

c. Reference climatological stations

There were 10 points available for (number of years of data as a RCS)/50.

RCSs were selected by countries as their best climate stations (World Meteorological Organization 1989a). Homogeneity was one of the criteria that went into their selection. Unfortunately, not all countries have sent the WMO a list of the RCSs, and how they were selected varied between countries. Therefore some weight was given to this aspect but not enough to prevent the selection of a very good nearby station from a country that had not yet provided its RCS list over an average RCS station.

d. Current reports

The maximum of either 10 points if data available 1990 or later, 20 points times the fraction of synoptic reports in 1995, 20 points times the fraction of CLIMAT reports in 1995, or 15 points if it was a RCS station.

To be most effective, all the GSN stations should be able to exchange their data in near-real time. The first factor considered is, are data present in GHCN or Jones for the station in 1990 or more recently? The purpose for this is essentially to give less weight to those stations for which no data have been received recently. The station would get additional weight if it reported regularly over the synoptic or CLIMAT network. Additionally, if a country thought highly enough of the station to designate it as an RCS, there was probably a better chance that its data might be exchanged in the future, hence some additional weight for the RCSs.

e. Population

The maximum of 20 points if rural, 15 points if a small town, 10 points if the population is unknown, and 0 points if urban.

Urban warming is a well-known phenomenon that the GSN would like to avoid. Therefore, more weight was given to stations that were rural or small town. While the local meteorological effect of urbanization is due to the land use and land cover (Gallo et al. 1996), population provides us with a useful, though less direct, criterion to assess urbanization. The population metadata were determined by locating the station on Operational Navigation Charts (ONCs). ONCs have been created by the U.S. Department of Defense and distributed by the National Oceanic and Atmospheric Administration. They are designed for pilots and, with a scale of 1:1 000 000, they provide well-defined topography, airport locations, and boundaries for urban areas. Once the station location is determined, if it is clearly not a rural site, the population of the town the station is associated with was determined using a variety of atlases. The station was designated as rural if there were less than 10 000 people in the community, small town if between 10 000 and 50 000, and urban if greater than 50 000.

These population metadata were developed for GHCN, so not all the source stations had population metadata. However, many additional station populations were determined for the GSN station selection process to make sure all selected stations and serious candidates for selection had population metadata. While these metadata may be the best population metadata currently available for the entire globe, some of the ONCs are a decade or two old. Therefore, some of the rural–urban boundaries are no longer accurate and a few of the stations should be reclassified.

f. Other networks

A number of networks already exist. Stations were selected for these networks for a variety of reasons, some of which, like data quality, are GSN concerns. One criterion in common with most of these networks is the impact a network designation may have on the future of the station: if several stations in a region were being closed, it is probable that the station with special network designation would be more likely to be

maintained. Indeed, this is one of the reasons for the creation of the GSN. Therefore some weight, albeit a rather small weight, was given to each of the following networks. The goal was to give some additional weight if a station was also a GUAN station, for instance, but not let four such designations outweigh climatological factors, hence the small point values.

- If the station is in the Regional Basic Synoptic Network (RBSN), it receives two points.
- Global Atmospheric Watch (GAW; World Meteorological Organization 1989b) stations receive one additional point.
- GUAN stations earn one point.
- In addition to having data for 1961–90 factored in, a station receives four points if it is a WMO normals station (World Meteorological Organization 1996b).
- Agrometeorological (AGRIMET; World Meteorological Organization 1989a) stations get two points.

Summing up all the potential points produces a total of 100. There were about 20 stations from GHCN whose data ended prior to 1896 and received zero points. The highest scoring station was Valentia Observatory in Ireland with 97.9 points. The median score was 47.

5. Selecting the network

Once the stations were rated on quality from a GSN perspective using the above algorithm, the best station in each area needed to be selected. There are many possible approaches for determining the geographic distribution of the stations. One would be to have a greater density of stations in regions that observations or general circulation model runs indicate have the greatest climatic variability or potential for climate change (e.g., Madden and Meehl 1993). Unfortunately, a spatial selection designed to optimize specific research would likely diminish the network's usefulness for different research. Originally, the participants in the Joint CCL–CBS Expert Meeting on the GCOS Surface Network decided to select the best station in every 5° × 5° box where stations were available. This scheme had three desirable characteristics:

- the scheme is very clear,
- the scheme creates an increasing density of stations from the equator toward the poles (required because the spatial correlation of surface temperature is smaller in high latitudes than in low latitudes), and
- the scheme allows for simple analysis methodology.

When examining the results, however, some disadvantages also became apparent: The spatial distribution was uneven and the scores were lower than desired. For these reasons an alternative approach was developed. The basic idea in this approach was as follows. First, a minimum required distance, R, between stations is defined. Then the highest ranking station is selected and all stations closer to that station than distance R are deleted from the list. This process is repeated for the next highest ranking station and on down the list until all stations are either selected or deleted.

In order to have the same increase of station density from the equator to the poles, the radius R was made dependent on the latitude. In the boxes system, the area that is represented by one station decreases with a rate of the cosine of the latitude. Therefore, distances between stations should drop at a rate of the square root of the cosine of the latitude:

$$R = R_e\,[\cos(\text{lat})]^{1/2},$$

where R_e represents the minimum distance at the equator. However, from 60° latitude toward the pole, R was kept at a constant $R_e(0.5)^{1/2}$.

Application of this algorithm revealed the following characteristics of the resulting network.

1) The numbers of stations in different areas of the world and, in particular, in different latitude zones were about proportional to the numbers that were attained in the grid boxes system, except for the Southern Hemisphere between 30° and 60°. In this area, the alternative approach selected fewer stations. This feature could be attributed to the lack of large land areas.
2) The average quality of stations selected was lower than for the boxes system (65.4 vs 66.3).

The following two measures were taken to correct these problems.

1) The decrease rate for the distance in the Southern Hemisphere was enhanced by using

$$R = R_e \cos(\text{lat}),$$

which appeared to be effective for attaining com-

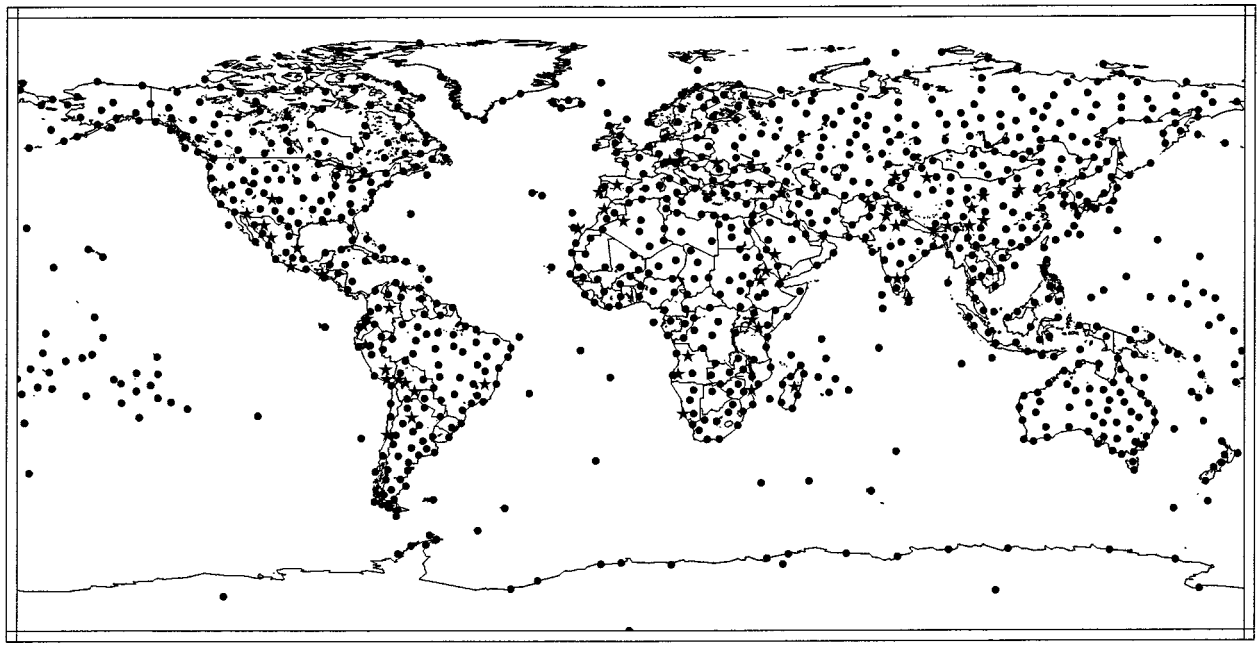

FIG. 3. The GCOS Surface Network: 940 stations are depicted with circles and the 60 stations selected to represent significantly different elevations are represented with stars.

parable numbers in all areas of the world (setting the equator distance R_e to 380 km).

2) The minimum distance was lowered to a figure allowing for about 1400 stations to be selected. From these 1400 stations, 460 stations were eliminated subjectively, region by region, with a view to optimal average score, but preventing gaps in the network that are too big. This resulted in a 940-station network, the same number of stations as the grid-box approach, but with an average rating of 68.2. The differences between the networks based on the grid boxes concept and the distance concept can be summarized as follows.

- The boxes system still has the advantage of its clarity; making changes when ratings may change is easier.
- A latitude-dependent density is incorporated in both systems.
- The distance-based network provides a better spatial distribution.
- The distance-based network provides a higher average quality.
- The distance-based network may require more sophisticated analysis techniques; such techniques, however, are not unusual in meteorology.

It was also decided at the Joint CCL–CBS Expert Meeting on the GCOS Surface Network that in mountainous regions, one station, particularly a valley floor station, could not adequately represent the climate of the region. Therefore, the task was to select, if possible, some additional stations at significantly different elevations in the areas of each continent with high standard deviations of elevation based on gridded elevation data (Row and Hastings 1994). The introduction of additional stations in these regions was done by introducing a vertical component in the concept of distance. This vertical component was defined by multiplying the difference in elevation between two stations with a factor of 300; that is, a difference in elevation of 1000 m has the same effect as a horizontal distance of 300 km. This produces an average minimum difference in elevation selection criterion of slightly over 1000 m. Sixty of these extra elevation stations were selected and their locations are indicated by stars in Fig. 3.

6. Results

With this change in the algorithm, the initial selection of the network was finalized, resulting in 1000 stations with an average quality of 67.6. The locations of the GSN stations are shown in Fig. 3. While there are fewer stations depicted in Fig. 3 than the 1634

CLIMAT reporting stations shown in Fig. 1, the spatial coverage of the GSN is more even. Only 55% of the initial selection of GSN stations currently send CLIMAT reports. The percentage of other metadata classifications are given in Table 1. Interestingly, while 50% of the source stations do not have WMO numbers, all but 8% of the stations selected for the GSN do. Since countries will typically only assign WMO numbers to stations whose data are being exchanged internationally, data for most of these stations are already being exchanged in some form. Therefore, as one might expect, most of these stations (96%) are also in the GHCN or Jones datasets.

Analysis of Table 1 reveals that the selected stations tend to have longer periods of record, are more rural, and are more likely to be designated RCSs. Indeed, over half the GSN stations are RCSs. This is not unexpected, since these are the features one would generally think of when looking for the "best" climate stations. This occurs even though selecting only a few stations in regions like the eastern United States and Europe, where there are many long-term stations available, while including most Antarctic stations, despite their short periods of record, decreases the potential mean period of record. Interestingly, there is a very high percentage of the selected stations that currently report synoptically (89%) and are part of the Regional Basic Synoptic Network (80%). This is because the weighting scheme gives some extra points to stations that are currently reporting. Over 4000 of the source stations reported synoptic data in 1995. Also, the vast majority (97%) of the CLIMAT reporting source stations also report synoptic data.

In addition to producing an initial selection of GSN stations, another result of this effort has been the production of valuable metadata (e.g., the rural/urban indicators and the links to other networks) about the surface stations that are widely used for climate research. These metadata will also allow individual countries to objectively compare the selection process to see why one station was selected for the GSN over a nearby station.

7. Discussion

The initial selection of the GSN was just the first step in a long process. The next major step was sending the list of selected stations, along with a list of the metadata obtained for all the stations, to the permanent representative of each WMO member country for evaluation. This will undoubtably result in some corrections to the metadata and changes in the final selection of the GSN. The very high percentage of selected stations that have WMO numbers and are in either the GHCN or Jones datasets implies either that most of the best stations are already exchanged internationally or that our knowledge of the other stations that might be available is limited. The evaluation by the permanent representatives will allow countries to compensate for the latter implication. It is also hoped that this effort will act as a catalyst for countries to review, update, and maintain adequate national RCS networks.

Although most of the GSN stations are already exchanged internationally, the type of international exchange is very important for climate research and monitoring. Because of the problems with missing observations or transmission errors, monthly means

TABLE 1. Station metadata classifications.

Station type	**GCOS surface network** (1000 stations)	**Source stations** (8653)
Without WMO numbers	8%	50%
In GHCN or Jones	96%	87%
Synoptic stations	89%	48%
CLIMAT stations	55%	18%
Reference climate stations	51%	26%
WMO normals station	63%	37%
AGRIMET stations	1%	2%
GAW stations	8%	2%
GUAN stations	9%	2%
RBSN stations	80%	33%
Rural stations	64%	54%*
Small town stations	18%	19%*
Urban stations	18%	27%*
Mean year station data start	1918	1926
Mean years of data since 1896	63	50

*For comparison purposes, these percentages were calculated from the total number of stations with population metadata. Of the 8653 stations in the source datasets, 12% did not have population metadata.

derived from transmitted daily synoptic reports will often have significant errors and biases when compared to monthly summaries calculated on site from the full month of data (Schneider 1992). Therefore, synoptic transmissions are not an adequate substitute for monthly CLIMAT style reports. Hopefully, adding one CLIMAT monthly transmission should be fairly easy for stations that regularly send out synoptic reports. Since 89% of the selected stations reported synoptic data in 1995, it is not unreasonable to expect that, with some encouragement, the WMO member countries will eventually report monthly climate summaries from all GCOS Surface Network stations.

Acknowledgments. We would like to thank Peter Scholefield and Valerie Gerard of WMO for their assistance. This work has been partially supported by the NOAA Climate and Global Change Program on Climate Change Data and Detection and the U.S. National Climatic Data Center. The development of the Jones database has been supported since the early 1980s by the U.S. Department of Energy. It is currently supported by Grant No. DE=FG02-86ER60397 from the Atmospheric and Climate Research Division.

References

Easterling, D. R., T. C. Peterson, and T. R. Karl, 1996: On the development and use of homogenized climate datasets. *J. Climate,* **9,** 1429–1434.

Gallo, K. P., D. R. Easterling, and T. C. Peterson, 1996: The influence of land use/land cover on climatological values of the diurnal temperature range. *J. Climate,* **9,** 2941–2944.

Jones, P. D., 1994: Hemispheric surface air temperature variations: A reanalysis and an update to 1993. *J. Climate,* **7,** 1794–1802.

Madden, R. A., and G. A. Meehl, 1993: Detecting greenhouse warming with the current surface network. *J. Climate,* **6,** 2486–2489.

Nowlin, W. D., Jr., and Coauthors, 1996: An ocean observing system for climate. *Bull. Amer. Meteor. Soc.*, **77,** 2243–2273.

Obasi, G. O. P., 1992: Letter, 29 October 1992. World Meteorological Organization No. M/CLC.

Peterson, T. C., and Vose, R. S., 1997: An overview of the Global Historical Climatology Network Temperature Database. *Bull. Amer. Meteor. Soc.*, in press.

Row, L. W., III, and D. A. Hastings, 1994: *TerrainBase Worldwide Digital Terrain Data, Documentation Manual and CD-ROM.* National Geophysical Data Center, Boulder, Colorado.

Schneider, U., 1992: The GPCC quality-control system for gauge-measured precipitation data. Report of a GEWEX Workshop, WMO/TD-No. 558, 153 pp. [Available from the World Meteorological Organization, Case Postal No. 2300, CH-1211, Geneva 2, Switzerland.]

Spence, T., and J. Townshend, 1995: The Global Climate Observing System (GCOS). An editorial. *Climate Change,* **31,** 131–134.

Vose, R. S., R. L. Schmoyer, P. M. Steurer, T. C. Peterson, R. Heim, T. R. Karl, and J. Eischeid, 1992: The global historical climatology network: Long-term monthly temperature, precipitation, sea level pressure, and station pressure data. ORNL/CDIAC-53, NDP-041, 296 pp. [Available from Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, TN 37831.]

World Meteorological Organization, 1988a: WMO technical regulations, general meteorological standards and recommended practices. Pub. 49, loose-leaf.

——, 1988b: WMO technical regulations, annex II: Manual on codes. Pub. 6, loose-leaf.

——, 1989a: Guide on the global observing system. Pub. 488, loose-leaf.

——, 1989b: The global atmosphere watch. World Meteorological Organization Fact Sheet 3, 4 pp.

——, 1994: Report of the GCOS Atmospheric Observation Panel, First Session, Hamburg, Germany. WMO-TD/No. 640, 16 pp.

——, 1995: Plan for the Global Climate Observing System (GCOS), WMO/TD No. 681, 49 pp.

——, 1996a: Weather reporting, observing stations. Pub. 9, loose-leaf.

——, 1996b: Climatological normals (CLINO) for the period 1961–1990. WMO/OMM No. 847, 768 pp.